The synchronization adjusting unit **14** receives the processing results from the prosody processing unit **13**, and adjusts the time durations for every phoneme to synchronize the image signal by using the synchronization information which was received from the multi-media distributor **11**. With the adjustment of the time duration of phonemes, the lip shape can be allocated to each phoneme in accordance with the position and manner of articulation for each phoneme, and the series of phonemes is divided into small groups corresponding to the number of the lip shapes recorded in the synchronization information by comparing the lip shape allocated to each phoneme with the lip shape in the synchronization information.

The time durations of the phonemes in each small group are calculated again by using information on the time durations of the lip shapes which is included in the synchronization information. The adjusted time duration information is made to be included in the results of the prosody processing unit **13**, and is transferred to the signal processing unit **15**.

The signal processing unit **15** receives the processing results from the synchronization adjusting unit **14**, and generates a synthesized speech by using the synthesis unit DB **16** to output it. The synthesis unit DB **16** selects the synthesis units required for synthesis in accordance with the request from the signal processing unit **15**, and transfers required data to the signal processing unit **15**.

In accordance with the present invention, a synthesized speech can be synchronized with moving picture by using the method wherein the real speech data and the shape of a lip in the moving picture are analyzed, and information on the estimated lip shape and text information are directly used in generating the synthesized speech. Accordingly, the dubbing of target language can be performed onto movies in foreign languages. Further, the present invention can be used in various applications such as a communication service, office automation, education, etc. since the synchronization of image information with the TTS is made possible in the multi-media environment.

The present invention has been described with reference to a particular embodiment in connection with a particular application. Those having ordinary skill in the art and access to the teachings of the present invention will recognize additional modifications and applications within the scope thereof.

It is therefore intended by the appended claims to cover any and all such applications, modifications, and embodiments within the scope of the present invention.

What is claimed is:

1. A system for synchronization between a moving picture and a text-to-speech converter, comprising:

distributing means for receiving multi-media input information, transforming said multi-media input information into respective data structures, and distributing the respective data structures for further processing;

image output means for receiving image information of the distributed multi-media information and displaying the image information;

language processing means for receiving language texts of the distributed multi-media information, transforming the language texts into phoneme strings, and estimating and symbolizing prosodic information from the language texts;

prosody processing means for receiving the prosodic information from said language processing means, and calculating values of prosodic control parameters;

synchronization adjusting means for receiving the prosodic control parameters from said prosody processing means, adjusting time durations for every phoneme for synchronization with the image information by using synchronization information of the distributed multi-media information, and inserting adjusted time durations into the prosodic control parameters;

signal processing means for receiving the processing results from said synchronization adjusting means and generating a synthesized speech; and

a synthesis unit database block for selecting required units for synthesis in accordance with a request from said signal processing means, and transmitting the required data to said signal processing means.

2. The system according to claim **1**, wherein the multi-media information comprises:

the language texts, image information on moving picture, and synchronization information,

and wherein the synchronization information includes:

a text, information on a lip shape, information on image positions in the moving picture, and information on time durations.

3. The system according to claim **2**, wherein the information on the lip shape can be transformed into numerical values based on a degree of a down motion of a lower lip, up and down motion at a left edge of an upper lip, up and down motion at a right edge of the upper lip, up and down motion at a left edge of the lower lip, up and down motion at a right edge of the lower lip, up and down motion at a center portion of the upper lip, up and down motion at a center portion of the lower lip, a degree of protrusion of the upper lip, a degree of protrusion of the lower lip, a distance from the center of the lip to the right edge of the lip, and a distance from the center of the lip to the left edge of the lip,

and wherein the information on the lip shape is definable in a quantified and normalized pattern in accordance with the position and manner of articulation for each phoneme.

4. The system according to claim **1**, wherein said synchronization adjusting means comprises means for calculating time durations of phonemes within a text by using the synchronization information in accordance with a predicted lip shape determined by a position and manner of articulation for each phoneme within a text, a lip shape within the synchronization information, and time durations.

* * * * *